Building Extraction and Change Detection in Multitemporal Aerial and Satellite Images in a Joint Stochastic Approach

Csaba Benedek, Xavier Descombes, Josiane Zerubia

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Building Extraction and Change Detection in Multitemporal Aerial and Satellite Images in a Joint Stochastic Approach

Csaba Benedek — Xavier Descombes — Josiane Zerubia

N° 7143

December 2009

Thème COM

ISSN 0249-6399 ISRN INRIA/RR--7143--FR+ENG

Building Extraction and Change Detection in Multitemporal Aerial and Satellite Images in a Joint Stochastic Approach

Csaba Benedek* , Xavier Descombes* , Josiane Zerubia*

Thème COM — Systèmes communicants
Projets ARIANA

Rapport de recherche n° 7143 — December 2009 — 40 pages

Abstract: In this report we introduce a new probabilistic method which integrates building extraction with change detection in remotely sensed image pairs. A global optimization process attempts to find the optimal configuration of buildings, considering the observed data, prior knowledge, and interactions between the neighboring building parts. The accuracy is ensured by a Bayesian object model verification, meanwhile the computational cost is significantly decreased by a non-uniform stochastic object birth process, which proposes relevant objects with higher probability based on low-level image features.

Key-words: Change detection, building extraction, marked point process, birth and death dynamics

* Ariana Project Team (INRIA/CNRS/UNSA), Sophia Antipolis, France

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Extraction de bâtiments et détection de changements sur des images aériennes et satellitales par une approche stochastique

Résumé : Dans ce rapport, nous proposons une nouvelle méthode probabiliste qui intègre l'extraction de bâtiments et la détection de changements à partir de paires d'images de télédétection. Un algorithme d'optimisation globale permet de trouver la configuration optimale de bâtiments en considérant des observations, des connaissances a priori et des interactions entre des parties voisines de bâtiments. La précision est assurée par une vérification d'un modèle objet bayésien; le coût du calcul est considérablement réduit en utilisant un processus stochastique non-uniforme de naissance d'objets fondé sur des caractéristiques bas-niveaux des images, qui génère des objets pertinents ayant une grande probabilité.

Mots-clés : Détection de changements, extraction de bâtiments, processus ponctuels marqués, dynamique de naissance/mort

Contents

1 Introduction

Remote sensing image analysis is a growing field of interest as the amount and quality of the available images, as well the number of related applications are rapidly increasing. In the recent years, the spatial resolution of the airborne and satellite image sensors has drastically been improved, providing more accurate and detailed information for the remote sensing image databases. This improvement enables to describe the Earth surface not only at the region level, but also at the object level: buildings, individual trees, can be extracted, leading to an increasing number of new applications, in urban development or forest monitoring, change detection or geographic information system databases.

Following the evolution of built-up regions is a key issue of the above domain, and has a vast bibliography. Numerous methods address building extraction at a single time instance [1, 2, 3]. The different approaches show a wide variety depending on the type, quality and number of the input images. It is common to use multiview inputs [4, 5] to exploit 3-D information in building modeling. The detection can be significantly facilitated by working on stereo-based Digital Elevation/Surface Models (DEM/DSM), where the silhouettes of the building footprints can be separated from the ground planes by the estimated height data [2, 6, 7, 8]. Another benefits are provided by multiple sensor inputs. Using Color InfraRed (CIR) images [9], the healthy vegetation can been identified by finding peaks in the near infrared wavelength band, thus significant image parts can be excluded from the building detection process. Further possibility is the fusion of aerial images with laser data [10] which directly provides surface information, similarly to the 3-D approaches. However several image repositories lack of stereo or multi-sensor information. This case is addressed in this report as well, thus building identification becomes here a challenging monocular object recognition task based on purely optical data [11].

As up-to-date remote sensing image databases contain often multitemporal image samples from the same geographical areas, change recognition and classification play currently a crucial role in the applications. Several recent building change detection approaches [7, 12] assume that for the earlier time layer a topographic building database is already available, thus the process can be decomposed into old model verification and new building exploration phases. On the other hand, many image repositories do not contain meta data, therefore the task requires automatic building detection in each image.

An object oriented change detection approach is introduced in [13] and applied for the extraction of damaged buildings after a Tsunami disaster. This method uses independent building detection processes in the two images which is followed by object level comparison. The later step is based on matching the geometry and spectral characteristic of the corresponding building candidates in the two time layers. However the object detection phase can be corrupted by image noise, irregular structures or occlusion by vegetation [13] which may present missing or only partially extracted buildings to the object matching module. Moreover the comparison may be affected by further intensity artifacts caused by shadows or altered illumination conditions.

Several low level change detection methods have been proposed for remote sensing [14, 15], which search for statistically unusual differences between the images without using

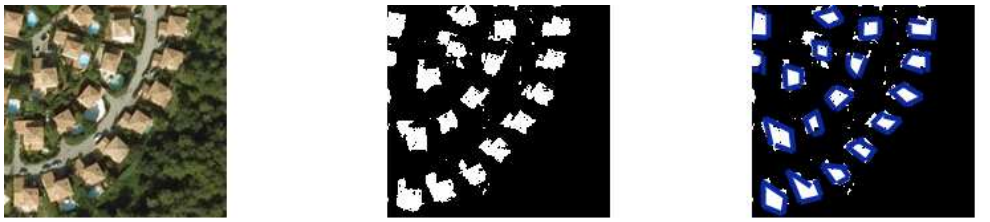


Figure 1: Building extraction from a single image (in the left) with a simple *bottom-up* method. The binary roof mask (in the middle) is obtained by color filtering, which is followed by connected component extraction (CCE) and polygon fitting steps (right).

explicit object models. These method must consider that the addressed photos are usually taken within a time interval of several years at different seasons and with different lighting conditions. In this case, simple techniques like thresholding the difference image [16,17, Sec. IV.] or background modeling [18] cannot be adopted efficiently since the observed pixel levels may be significantly different even in the 'unchanged' image regions.

The change detection algorithms in the literature follow either the Post-Classification Comparison (PCC) or the direct approach. *PCC* models [19, 20, 21, 22, 23, 24] segment the input images with different land-cover classes, like arboreous lands, barren lands and artificial structures [22]. Thus changes are obtained indirectly here as regions with different classes in the two image layers. On the other hand *direct* methods [25, 26, 27, 28] derive a similarity-feature map from the input photos [e.g. a Difference Image (DI)] and then segment the feature map to separate changed and unchanged areas. As for PCC approaches, besides change detection, they classify the observed differences at the same time (e.g. a barren land turns into a built-up area); and the quality of their results can be enhanced by interactive segmentation of the images [23] or exploiting estimated class transition probabilities [22]. However using PCC models we have to fix the clusters a priori in the scenes, and we need to find reliable feature models for each land-cover class with probably various subclasses. Moreover, in object-focused applications 'intra-class' transitions - which are ignored by PCC methods - may also be worth of the attention: e.g. in our case it is necessary to detect destroyed, modified or newly built buildings inside an urban region.

On the other hand, most change detection methods are based on the assumption that changes occur very rarely, therefore in a given image the area of the changed regions is statistically negligible compared to the unchanged territories (i.e. *background*). In these cases the global statistical properties (e.g. histogram) of the features over the whole image approximate the feature statistics over the unchanged regions, thus after background model estimation the changes can be identified by outlier detection [29, 27]. However, in dynamically improving (sub-)urban areas this assumption is often invalid (see later in Fig. 13 and 14 image pairs), calling for solutions which are insensitive to the quantity of differences.

Figure 2: Demonstration of inverse techniques. The fitness function can be evaluated for each possible object configuration, however, it must result in a low fitness value for incorrect populations (left), and a high fitness value for the relevant ones (right)

Although the above detailed low level change detection algorithms are usually considered as preprocessing filters, there have been less attempts given to justify how they can support the object level investigations. We try to step forward in this report, and combine object extraction with local low level similarity information between the corresponding image parts in a unified probabilistic model. It will be shown that we can benefit from evidences such as building changes can be found in the *changed* areas, while multiple object views from the different time layers may increase the detection accuracy of the *unchanged* buildings.

Another important issue is related to object modeling. The *bottom-up* techniques [1] construct the buildings from primitives, like roof blobs, edge parts or corners. Fig. 1 shows an example: first a binary roof map is extracted by a simple color filter, thereafter polygons are fitted to the borders of the large connected components. Although the bottom-up methods can be fast, they may fail if the primitives cannot be reliably detected.

To increase robustness, it is common to follow the Hypothesis Generation-Acceptance (HGA) schema [3,30]. Here the accuracy of object proposition is not crucial, as false candidates can be eliminated in the verification step. However objects missed by the generation process cannot be recovered later, which may result in several missing alarms. If too many object hypothesises should be checked (e.g. applying exhaustive search) the detection process can be unfeasibly slow. Finally, these techniques search for separate objects instead of global populations, disregarding population-level features such as overlapping, relative alignment, color similarity or spatial distance of the neighboring objects [2].

To overcome the above defects, recently proposed *inverse methods* [31] assign a fitness value to each possible object configuration and an optimization process attempts to find the configuration with the highest confidence (see Fig. 2). In this way, flexible object appearance models can be adopted, and it is also straightforward to incorporate prior shape information and object interactions. However, to keep the computational tractability, this approach needs to perform efficient searching in the high dimension population space, where local maxima of the fitness function can mislead the optimization.

Since the seminal work of Geman and Geman [32] inverse techniques have been extensively used to solve various image processing problems. Markov random fields (MRFs) proved to be efficient among others in different classification tasks ensuring smooth and observation-consistent segmentation at the same time. However object information is only very partially contained in the pixel radiometry or texture, which is unable to appropriately address the geometric content of the image. Considering this demand, the conventional MRF frameworks have been recently extended [31], by taking into account the geometry in the proposed models. To really take advantage of the very high resolution, it seems to be efficient to work with objects as variables rather than with pixels. In such a case, the number of variables (number of objects) is also unknown. Marked point processes (MPP) [31] are good candidates to address both challenges: model the geometry of objects and work with an unknown number of variables. Moreover, they can also embed prior constraints and data models within the same density, similarly to MRFs. MPP models became recently well established in several applications, which aim at detecting multiple but in some sense similar objects; like trees in forests or plantations [33, 34, 35], flamingos in a colony [36, 37], buildings in city DEM maps [2,6,38,39,40], or stereo images [41, 42], line network extraction [43], or addressing general feature detection tasks [44].

In this report, we propose a novel MPP approach for the building change detection problem, devoting special attention to the optimization issue. We attempt to merge the advantages of both low level and object level approaches in a consistent probabilistic framework. The applied Multiple Birth and Death technique [45] evolves the population of buildings by alternating object proposition (*birth*) and removal (*death*) steps. The exploration in the configuration space is driven by simple pixel and region descriptors, however the object verification follows the robust *inverse* modeling approach. Unlike in conventional HGA techniques, stochastic processes are used here for both object proposition and acceptance, while a simulated annealing framework ensures convergence of the dynamics. Due to the high modularity, the proposed model could be easily adapted to different object level change detection applications, for example tree, road or river detection. On the other hand, we attempt to focus on some task specific issues as well. We present a wide feature library, which can be appropriate for the identification of a large set of buildings, expecting various image properties. For this reason, we give in the following section an overview on the state-of-the art methods of building extraction.

1.1 State of the art in single-view building detection

A SIFT keypoint based approach has been introduced in [46] for urban area extraction and building detection. This method assumes that the building structures in a given image can be efficiently characterized by a couple of template buildings (here two templates: a bright and a dark one) which are used for training. The goal is localization, but the accurate bounding boxes of the buildings are not extracted which makes difficult to apply the method for change detection. As well, images containing a high variety of buildings may need a huge template library, where the overlap between the buildings and background in the descriptor domain can be hardly controlled.

A stochastic framework is presented in [1] for detecting building rooftops from single images, which combines 2-D and 3-D information. This approach is based on hierarchical grouping of extracted edge segments to form continuous lines, junctions and finally closed curve hypothesizes. However, several restrictions are used for the buildings, assuming that they have uniform height, they are composed of planar surfaces with parallel sides and each building casts its shadow on a locally flat surface. As well, similarly to [30, 47], the method needs a reasonable edge map, because missing large side parts, or plenty of false edges inside and around the buildings may corrupt the edge grouping process.

Following a different approach from edge based techniques, building detection is often considered as a region level or image segmentation problem [48, 49, 50]. In [50] the authors assume that buildings are homogenous areas either in color or in texture, which can be used for training-based background subtraction. Thereafter elementary constraints for shape and size are used to group the candidate regions into building objects. This method can fail, if due to the weak contrast several building and background parts are merged in the same region of the oversegmented map, or the background and building areas are strongly overlapped in the chosen feature domain. On the other hand, in case of homogenous building appearances (see BEIJING, Fig. 24, bottom) or presence of flagrant roof colors (see BUDAPEST red roofs, Fig. 24, top) region features may be more robust than weak or ragged edge maps – however we must expect that this approach will retrieve only a part of the real objects [3].

The common property of the previous techniques is that they are based on one or more specific hypothesizes (like presence of unique roof colors, shadows and shadow filters, strong edges, homogenous roofs, only a few typical building structures, or simple 3-D models can be fit), but they fail if the used features are missing or less discriminative in the input data. Apart from a few models it is not straightforward how to adopt a method depending on different circumstances (an exception is [51] where data dependent model part can be exchanged without modifying the prior term). However to increase the generality and robustness, besides extracting the descriptors, feature integration and selection should be addressed at the same time. Therefore we aim to construct a framework which can combine the features in a flexible way based on availability, making them to work for an extended set of images and situations.

A two-step method for building extraction based on roof color, shadow and edge information has been proposed in [3]. First, red roofs and shadow regions are filtered in illumination invariant color spaces, and Candidates of Built-up Territories (CBT) are identified either as roof-colored image regions, or as areas located next to the shadow blobs in the estimated *sun direction*. Thereafter, houses are defined as rectangular structures inside the CBTs. The second step is responsible for verification and fitting the accurate building shapes to the object candidates purely based on (Canny) edge information in the CBTs. The key point of the applied box fitting algorithm is to determine a perpendicular junction in the contained line segments which fixes a corner and the orientation of the house rectangle. The remaining two parameters (side lengths) are set by exhaustive search, looking for the best match between the Canny edge map and estimated rectangular edge mask (using a Hausdroff-like distance). The quality of the edge mask is crucial for the process. Although in the verifi-

cation phase, the Hausdorf-matching is relatively robust, the corner detection and/or the orientation estimation can fail if the corners are weak or the edge map is strongly corrupted. As well the assumption that shadows can be filtered in the 'blue' color domain is often invalid, and in several cases we must expect that the candidate regions of the neighbouring houses often overlap. Although this method integrates three different features, color and shadow information are only used for roughly estimating the object locations, and the binary edge map is responsible also for corner detection, orientation proposition and building verification making the process sensitive to the edge extraction. If neither shadow nor color information is available the building search area should be extended to the whole image which can significantly increase the processing time, meanwhile other rectangular structures may be erroneously detected as buildings.

Beside probabilistic models [2, 6], variational techniques [51, 52] have been recently proposed for building extraction through global energy minimization processes. Similarly to our framework, [51] uses *data* and *prior* decomposition and handles these two issues separately in the modeling phase. Here they focus primarily on the prior model, and use only a simple data term which can be replaced later with different object appearance models independently from the priors.

From another point of view of prior shape modeling, some of the methods use libraries of complete 2-D [51] or 3-D [2] object shapes, while others [1, 6] construct the objects from elementary building stones (rectangles or line segments), and the higher level shape information is encoded by interaction constraints of the nearby components. While the global description can be efficient if the appearing building shapes can be characterized by a restricted number of prototypes, the constructive approach - which we follow in the current report as well - is more general if the prior models of the complete buildings are partially unknown or have a huge variety.

1.2 Shadow detection in aerial and satellite images

As underlined in the previous section, shadows are widely used in the building localization process. This step needs principally the extraction of the shadowed regions which is itself a hot topic of research [53] and has its own literature for remote sensing applications [54, 55, 56, 57, 58, 59].

The large variety of the input data calls for different approaches to tackle the problem. Techniques addressing High Resolution (HR) satellite imagery [56,59] usually deal with single channel images. Since here the only available pixel information is the intensity, one can – at pixel level – solely rely on the assumption that shadows correspond to dark image areas. It is also important to note that for building detection the cast shadows (i.e. shadows on the ground) are only relevant, while self shadows (i.e. weakly or not illuminated building parts) should be ignored. However, as pointed out in [56], in most cases cast and self shadows have different intensity values, since the shadowed object parts are illuminated more by secondary light sources such as reflection from surrounding buildings.

Performing shadow detection via pixel brightness filtering faces two main challenges. Firstly, we find usually a significant overlap between the intensity domains of the shadowed

and non-shadowed areas, therefore some misclassified regions are expected. Secondly, the separation – commonly thresholding [56]– must be appropriately parametrized. There have been a few methods proposed for automated threshold estimation, most frequently based on global image histograms. For example, in [56] the threshold is calculated as the mean of the two main peak locations, which can be appropriate for bi-modal histograms, but less efficient in images with uniform or strongly multi-modal intensity statistics, such as in the BUDAPEST photo pair (see Fig.24). Instead of the less general unsupervised approaches, it is often allowed to set the optimal threshold by user interaction or by training data, especially if the database has several image samples with the same quality parameters and illumination conditions.

On the other hand, the intensity domain overlapping problem cannot be eliminated at pixel level, but some attempts have been proposed to tackle it through region filtering. In [56], filtering is mainly based on simple region attributes (size and location) of the intensity-based shadow map. As the authors point out, there is often a huge overlap between the radiometric domains of shadow and water, similarly to the BEIJING images [see Fig. 11(b)]. They attempt to separate the watered regions as they are very homogenous (i.e. have low variance), while they assume that the shadowed areas are more textured (high variance). Unfortunately, we found in a few cases that due to sensor saturation, both the shadows and rivers consist of homogenous zero-pixel-valued regions, which make such a discrimination inaccurate.

Pixel-level shadow filtering techniques can benefit from color image inputs [54,55,57], exploiting photometric evidences, such as shadows cause increased hue values or higher saturation with short blue-violet wavelength [55]. These methods work usually in color spaces with separated luminance (intensity-equivalent) and chromaticity (hue-equivalent) channels, such as HSI, HSV, HCV etc. [55]. In general, color information projects the classification problem from the intensity domain to 2-D or 3-D vector spaces, which encapsulate more information for the separation, but often need the estimation of more parameters at the same time. On the contrary, [55] derives the ratio map of the hue-equivalent and intensity-equivalent channels, and performs the classification based on the 1-D ratio descriptors. According to the experiments, the intensity based shadow filtering may be notably improved considering chromaticity values, however, they can only decrease but not eliminate the problem of domain-overlapping, which still means challenges for higher level model elements.

2 Problem formulation

The input of the proposed method consists of two co-registered aerial or satellite images which are taken from the same area with several months or years time differences. Thus a single view is available at each time instance, and we cannot exploit additional meta-information such as maps or topographic building databases. We expect the presence of registration or parallax errors, but we assume that they only cause distortions of a few pixels. We assume that the projection of the buildings onto the image plane consists of one

Figure 3: Description of the rectangle parameters

or many rectangular building segments, which we aim to extract by the model described below.

As for the output, in each image we provide the size, position and orientation parameters of the detected building segments, meanwhile we also give some information such as objects are new, demolished, modified/rebuilt and unchanged.

Denote by S the common $S_W \times S_H$ pixel lattice of the input images and by $s \in S$ a single pixel. Let u be a building segment candidate. We consider the center of each building, $c = [c_x, c_y]$ as a point process in $[0, S_W] \times [0, S_H] \subset \mathbb{R}^2$, which can be projected to S by simple discretization: $c \to [\lfloor c_x \rfloor, \lfloor c_y \rfloor]$. For purposes of dealing with multiple time layers we assign to u an image index flag from the set $\Xi = \{1, 2, *\}$, where '$*$' indicates unchanged object (i.e. present in both images), while '1' and '2' correspond to building segments which appear *only* in the first *or* second image respectively. Let be rectangle $R_u \subset S$ the set of pixels corresponding to u, which can be described by the five rectangle parameters as shown in Fig. 3. In summary, an object u is characterized by the following attributes:

- $(c_x, c_y) \in \mathcal{C}_x \times \mathcal{C}_y$ center coordinates, where $\mathcal{C}_x \in [0, S_W]$, $\mathcal{C}_y \in [0, S_H]$
- $(e_L, e_l) \in \mathcal{E}_L \times \mathcal{E}_l$ side lengths, where $\mathcal{E}_L = \{L_{\min}, \ldots, L_{\max}\}$, $\mathcal{E}_l = \{l_{\min}, \ldots, l_{\max}\}$
- $\theta \in \mathcal{O}$ orientation, where $\mathcal{O} = [-90°, +90°]$
- $\xi \in \Xi$ image index flag, where $\Xi = \{1, 2, *\}$

For example, we will denote by $\theta(u)$ the orientation of the object u.

Based on the previous definitions, the set of all the possible object records – denoted by $\mathcal{H}$ – has the following form:

$$\mathcal{H} = \mathcal{C}_x \times \mathcal{C}_y \times \mathcal{E}_L \times \mathcal{E}_l \times \mathcal{O} \times \Xi$$

Using the previous data structure, the classification of the building segment $u \in \mathcal{H}$ is straightforward:

- u is *unchanged* iff $\xi(u) = *$

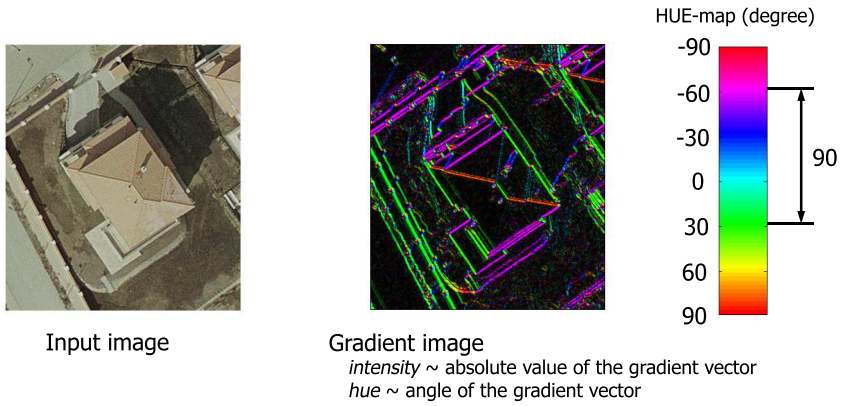


Figure 4: Gradient map of an image part - with visualizing both the magnitude (intensity) and orientation (hue) of the local gradient vectors.

- u is *new* iff $\xi(u) = 2$ and $\nexists v \in \omega : \{\xi(v) = 1, u \text{ and } v \text{ overlap}\}$.
- u is *demolished* iff $\xi(u) = 1$ and $\nexists v \in \omega : \{\xi(v) = 2, u \text{ and } v \text{ overlap}\}$.
- u is *modified/rebuilt*: otherwise

3 Feature selection

In the proposed model, low level and object level features are distinguished. Low level descriptors are extracted around each pixel such as typical color or texture, and local similarity between the time layers. They are used by the exploration process to estimate where the buildings *can* be located, and how they *can* look like: the *birth* step generates objects in the estimated built-up regions with higher probability. On the other hand, object level features characterize a given object candidate u, and are exploited for the fitness calculation of the proposed oriented rectangles. Building verification is primarily based on the object level features thus their accuracy is crucial. Since apart from the similarity measure, the upcoming descriptors are generated for the two input images separately, we often do not indicate the image index in this section.

3.1 Low level features of building identification

3.1.1 Local gradient orientation density

The first feature exploits the fact that building-regions should contain edges in *perpendicular* directions, which is demonstrated in Fig. 4. This property can be robustly char-

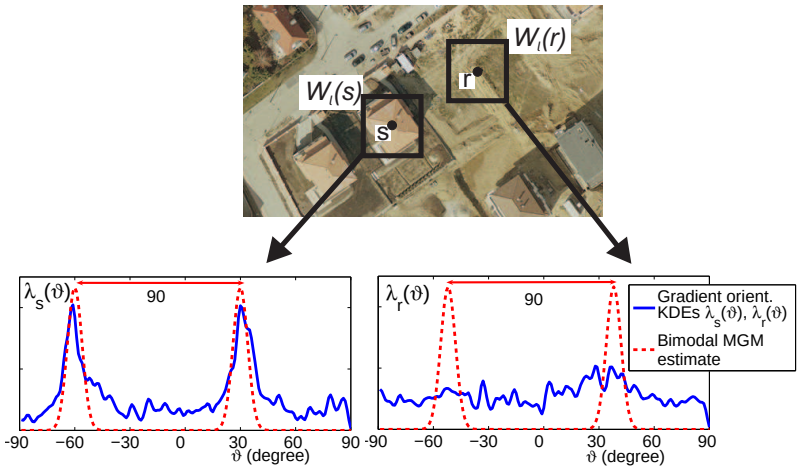


Figure 5: Kernel density estimation of the local gradient orientations over rectangles around two selected pixels: a building center s and an empty site r.

(a) Input image © András Görög (b) Detected building regions

Figure 6: Example of the thresholded P_b^α birth map for the BUDAPEST image

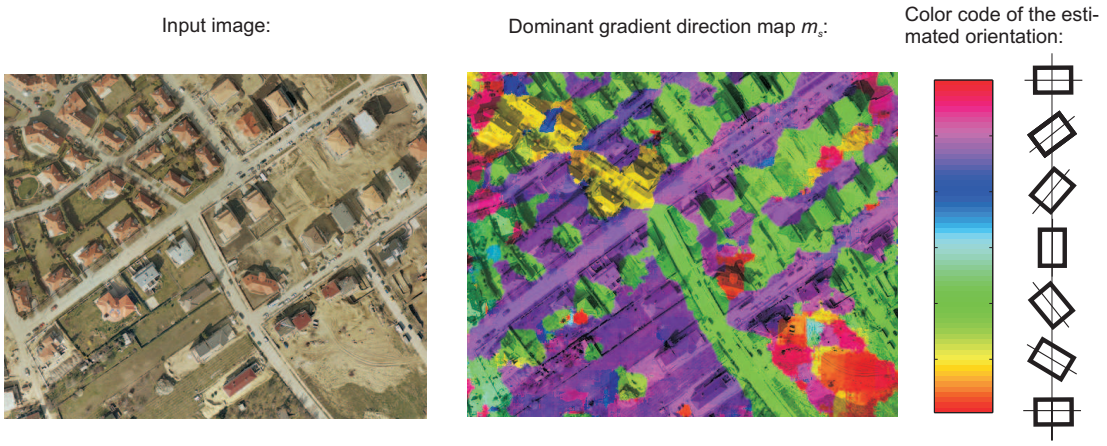


Figure 7: The $\{m_s|s \in S\}$ local dominant gradient direction map. Pixel colors (hue in the HSV color representation) correspond to different estimated orientations as coded in the right color bar, pixel intensities (value in HSV) are kept from the original image

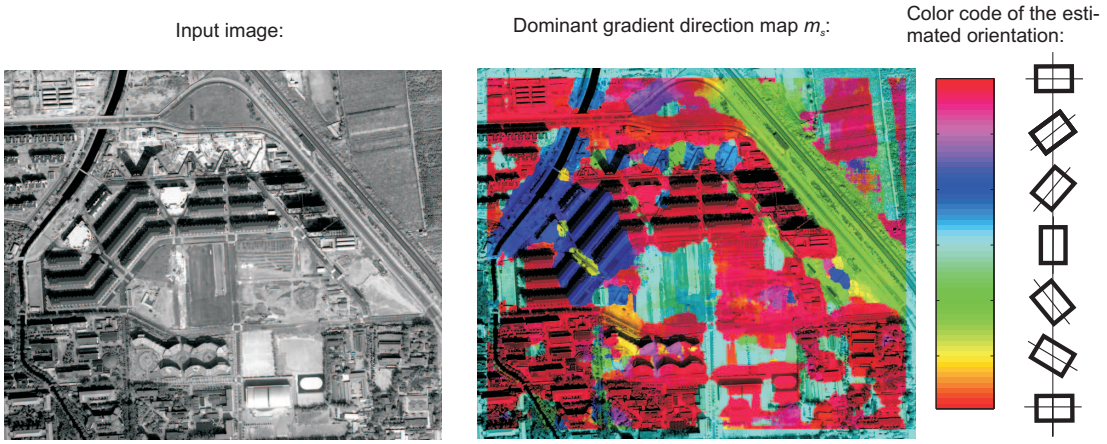
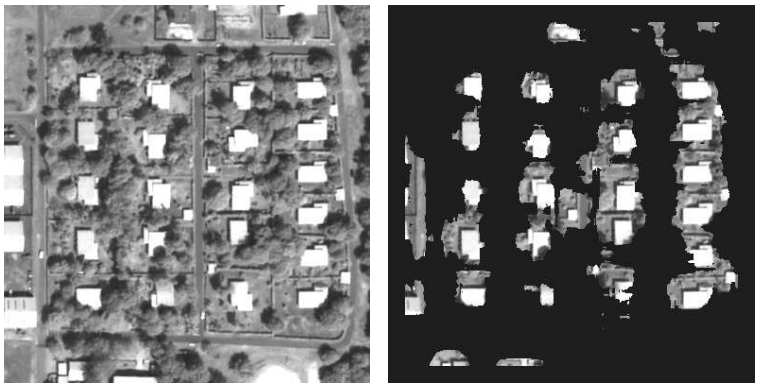


Figure 8: Demonstration of the $\{m_s|s \in S\}$ local dominant gradient direction map in the BEIJING image © LIAMA CAS. Different colors correspond to different estimated orientations as detailed on the right color bar



(a) Input image © DGA (b) Detected building regions

Figure 9: Demonstration of the thresholded P_s^α birth map for the ABIDJAN image

acterized by local gradient orientation histograms [60]. Let be $g = \{g_s | s \in S\}$ the intensity map and $\nabla g_s = [\nabla g_s^x, \nabla g_s^y]$ the intensity gradient vector at s with magnitude $||\nabla g_s|| = \sqrt{(\nabla g_s^x)^2 + (\nabla g_s^y)^2}$ and angle $\vartheta_s = \arctan(\nabla g_s^y / \nabla g_s^x)$. Let be $W_l(s)$ the rectangular $l \times l$ sized window around s, where l is chosen so that $W_l(s)$ can cover an average building with arbitrary orientation. For each s we calculate the weighted ϑ_s density of $W_l(s)$:

$$\lambda_s(\vartheta) = \frac{1}{N_s} \sum_{r \in W_l(s)} \frac{1}{h} \cdot ||\nabla g_r|| \cdot k\left(\frac{\vartheta - \vartheta_r}{h}\right)$$

where

$$N_s = \sum_{r \in W_l(s)} ||\nabla g_r||$$

and $k(.)$ is a kernel with bandwidth parameter h. We used uniform kernels for quick calculation. If $W_l(s)$ covers a building, the $\lambda_s(\vartheta)$ function has two peaks located at a distance of $90°$ in the ϑ-domain (see Fig. 5). This property can be measured by correlating $\lambda_s(\vartheta)$ with an appropriately matched bi-modal density function:

$$\alpha(s, m) = \int \lambda_s(\vartheta) \eta_2(\vartheta, m, d_\lambda)\, d\vartheta$$

where $\eta_2(.)$ is a mixture of two Gaussians with mean values m, resp. $m + 90°$, and deviation d_λ for both components (d_λ is a parameter of the process). Offset (m_s) and value (α_s) of the maximal correlation can be obtained as:

$$m_s = \text{argmax}_{m \in [-90°, 0]} \{\alpha(s, m)\} \qquad \alpha_s = \alpha(s, m_s)$$

Pixels with high α_s are more likely centers of buildings, which can be coded in an α-birth map $P_b^{\alpha}(s) = \alpha_s / \sum_{r \in S} \alpha_r$. The nomination comes from the fact that the frequency of proposing an object in s will be proportional to the local birth factor $P_b(s)$. The thresholded P_s^{α} maps are shown in Fig. 6 and 9 for two chosen images.

Moreover, offset m_s offers an estimate for the dominant gradient direction in $W_l(s)$ (see Fig. 7). Thus for object u proposed with center s, we model its orientation as $\theta(u) = m_s + \eta_s^{\theta}$, where η_s^{θ} is a zero-mean Gaussian random variable with a small deviation parameter σ_{θ}.

We have observed in various experiments (Fig. 6 and 9) that the α_s-gradient feature is usually able to roughly estimate the built-up regions. However, in several cases the detection can be refined considering other descriptors such as roof colors or shadows [3].

3.1.2 Roof color filtering

Some of the roof colors can be filtered using illumination invariant color representations, such as the a* channel in the CIE L*a*b* color space (see Fig. 10(c)). Assume that we can extract in this way a $\mu_c(s) \in \{0, 1\}$ indicator mask, where $\mu_c(s) = 1$ means that pixel s has roof color. We calculate the color feature for s as $\Gamma_s = \sum_{r \in W_l(s)} \mu_c(r)$ and the color birth-map as $P_b^c(s) = \Gamma_s / \sum_{r \in S} \Gamma_r$. Note that obviously this information cannot be used for grayscale inputs, and even in color images the $\mu_c(s)$ filter usually finds only part of the roofs which have typical 'red colors' ([3] and Fig. 10(d)).

3.1.3 Shadow evidence

As discussed in Sections 1.1 and 1.2, a supplementary evidence for the presence of buildings can be obtained by the detection of their cast shadows [1,3], exploiting that the darkness and direction of shadows are global image features. We have derived a (noisy) binary shadow mask $\mu_{\mathrm{sh}}(s)$ by intensity [56], resp. color filtering [55], techniques depending on the image input, as shown in Fig. 10(e). Thereafter building candidate regions can be identified as image areas lying next to the shadow blobs in the opposite shadow direction (see Fig. 10(f)). We used a constant birth rate $P_b^{\mathrm{sh}}(s) = p_0^{\mathrm{sh}}$ within the obtained candidate regions and a significantly smaller constant ϵ_0^{sh} outside.

3.1.4 Roof homogeneity

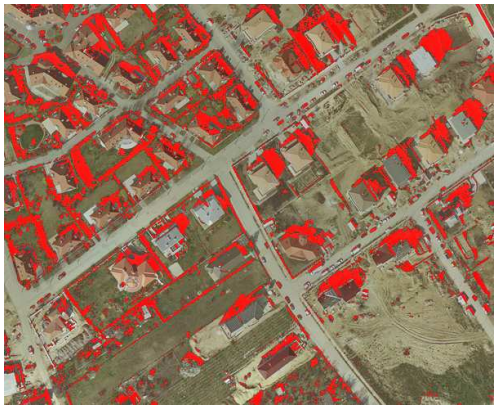
As shown previously, the $P_b^{\alpha}(s)$ and $P_b^{\mathrm{sh}}(s)$ birth maps give usually a quite coarse estimation of the built-up regions, which is hardly appropriate for connected component analysis based building separation. Although we may obtain notably accurate footprints through roof color filtering, it can be only used for a limited subset of the images and objects. Particularly in grayscale images, the overlap between the building and background intensity domains is usually too large for efficient separation. On the other hand, in images provided by HR satellites such as Ikonos or Quickbird, a significant part of the roof tops can be identified as homogenous blobs in the coarsely detected building candidate regions. In this section we investigate, how *roof homogeneity* can be exploited in the building region refinement process.

(a) Input image © András Görög

(b) Ground truth buildings (in red)

(c) The a* channel in the CIE L*a*b* color space

(d) Color based building area detection



(e) Detected shadows

(f) Shadow based building area detection

Figure 10: Example of the color and shadow features

The feature calculation process consists of the following steps (an example for BEIJING image is shown in Fig. 11):

- `Candidate Region Filtering`: for a given input image (Fig. 11(a)), obtain the coarse preliminary building candidate (PBC) regions based on the gradient and/or shadow features (Fig. 11(b)-(c)).
- `Intensity based segmentation`: we (over)segment the PBC regions of the input image into homogenous components (SPBC map, see Fig. 11(d)). Thereafter, we ignore the small blobs of SPBC to obtain the homogenous building candidate (HBC) region map (Fig. 11(e)). We performed the segmentation with a conventional floodfill propagation algorithm [61].
- `Orientation based clustering`: we re-cluster the HBC map based on the m_s dominant local gradient orientation values obtained in the regions of interest (see also Fig. 8 earlier), and call the result GHBC image as shown in Fig. 11(f). Each large connected component of GHBC is considered in the following as a building segment candidate.
- `Candidate parameter estimation`: we estimate the center (Fig. 11(g)) and the bounding box (Fig. 11(h)) parameters for each building segment candidate through simple morphological box fitting techniques.

Similarly to the color filter, the homogeneity feature can only describe a subset of the buildings: in Fig. 11 three houses on the right side and one at the bottom left are ignored since they have strongly textured roofs.

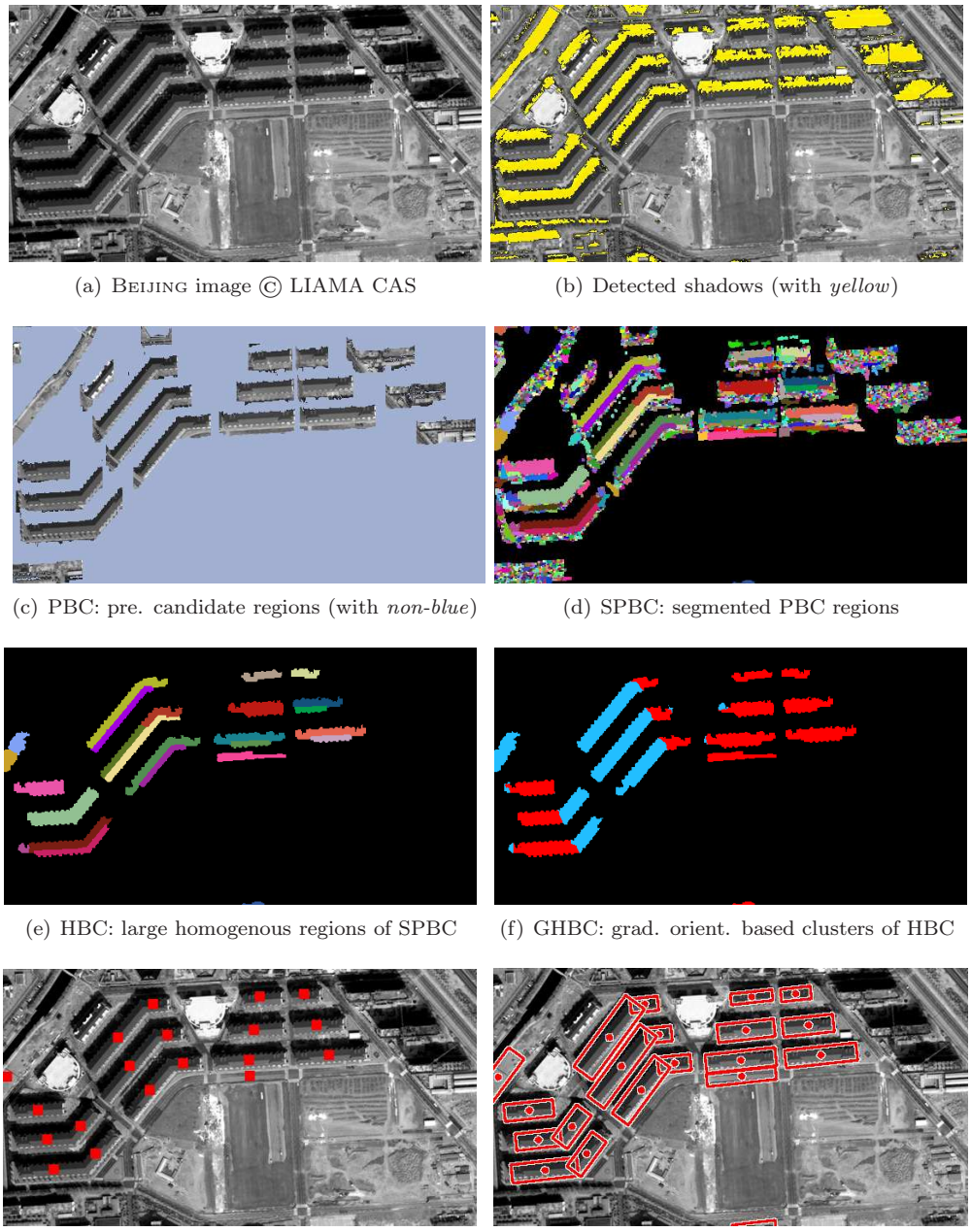
Denote the candidate rectangles obtained in the previous filtering process (Fig. 11(h)) by $\mathcal{R}_i$, $i = 1 \dots t$. The birth map is calculated as $P_b^h(s) = \max\{p_{\mathcal{R}}(s), \epsilon_{\mathcal{R}}\}$ where $\epsilon_{\mathcal{R}}$ is a small constant probability:

$$p_{\mathcal{R}}(s) = \sum_{i=t} k\left(\frac{||s - c(\mathcal{R}_i)||}{h_{\mathcal{R}}}\right)$$

with a $k(.)$ kernel function.

Besides marking the candidate regions of the building center, the $\{\mathcal{R}_i | i = 1 \dots t\}$ set provides estimation of the rectangle side length parameters (Fig. 11(h)). Of course, we can only keep this information reliable in the neighborhood of the homogenous building centers, thus first we generate a binary $\Upsilon_b(s)$ mask through thresholding the $P_b^h(s)$ map (if the homogeneity feature is ignored in the model we consider $\Upsilon_b(s)$ as a constant zero image). Thereafter, for each pixel s with $\Upsilon_b(s) = 1$ we find the closest rectangle $\mathcal{R}_s^{\min} = \arg\min_i ||s - c(\mathcal{R}_i)||$ and set $\mu_L(s) = e_L(\mathcal{R}_s^{\min})$ and $\mu_l(s) = e_l(\mathcal{R}_s^{\min})$ side length estimates. Later on, if an u object is proposed with the center at s and $\Upsilon_b(s) = 1$, we will choose its side length values as:

$$e_L(u) = \mu_L(s) + \eta_s^L$$



(a) Beijing image © LIAMA CAS (b) Detected shadows (with *yellow*)

(c) PBC: pre. candidate regions (with *non-blue*) (d) SPBC: segmented PBC regions

(e) HBC: large homogenous regions of SPBC (f) GHBC: grad. orient. based clusters of HBC

(g) Estimated building centers based on GHBC (h) Estimated building rectangles based on GHBC

Figure 11: Illustration of the homogeneity based pre-detection

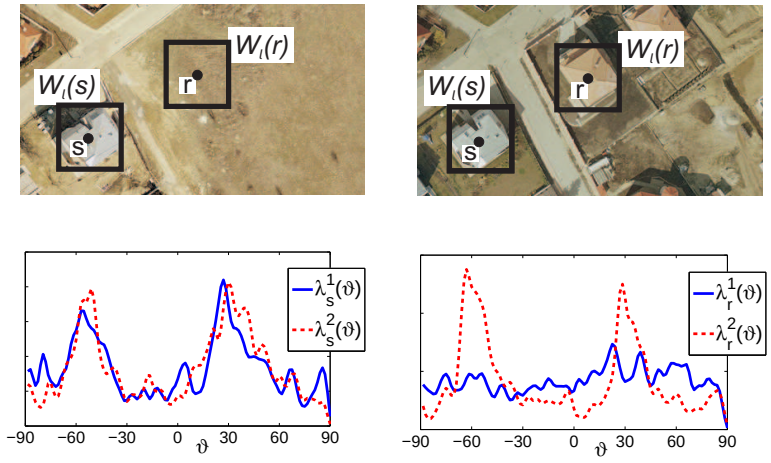


Figure 12: Comparing the $\lambda(.)$ functions in the two image layers regarding two selected pixels. s corresponds to an unchanged point and r to a built-up change.

$$e_l(u) = \mu_l(s) + \eta_s^l$$

where η_s^L and η_s^l are independent zero mean Gaussian random variables with standard deviation parameters σ_L and σ_l, respectively.

Note that the side length estimates can be similarly extracted from the color feature map. This preliminary calculation is particularly significant if the object sizes show a large variety as in the BEIJING images. Here sampling the parameters of the proposed objects according to a uniform distribution with wide support can critically slow down the speed of the evolution.

3.1.5 Integration of the different birth maps

Since the main goal of the *combined birth map* is to keep focus on all building candidate areas, we derive it with the maximum operator from the feature birth maps. For example, when gradient, color and shadow are simultaneously used, we obtain the final field as $P_b(s) = \max\left\{P_b^{\alpha}(s), P_b^{c}(s), P_b^{\mathrm{sh}}(s)\right\}\ \forall s \in S$. For input, without shadow or color information, we can ignore the corresponding feature in a straightforward way, or exchange the $P_b^c(s)$ color component to the birth value of the homogeneity feature, $P_b^h(s) \cdot \mathbb{I}_{\Upsilon(s)=1}$, where $\mathbb{I}$ refers to the indicator function. Note that we generate birth and orientation maps for both images which are denoted by $P_b^{(i)}(s)$, $m_s^{(i)}$, $i \in \{1, 2\}$.

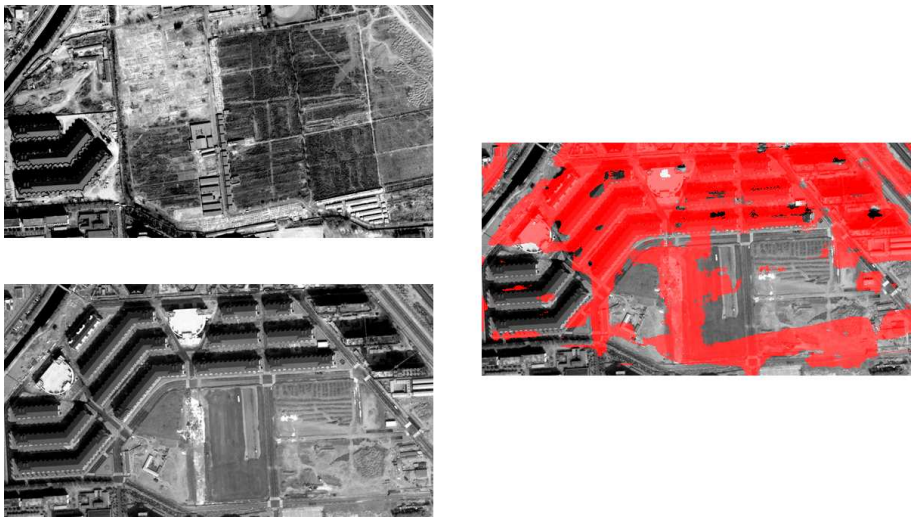


Figure 13: Detected changes with the correlation feature of the $\lambda(.)$ functions, BEIJING image pair © LIAMA Laboratory CAS

3.2 Low level similarity feature

The gradient orientation statistics also offers a tool for low level region comparison. Matching the $\lambda_s^1(.)$ and $\lambda_s^2(.)$ functions can be considered as low level similarity checking of the areas around s in the two images, based on "building-focused" textural features (see Fig 12), which are independent of illumination and coloring effects and robust regarding parallax and registration errors. For measuring the dissimilarities we used the Bhattacharyya distance:

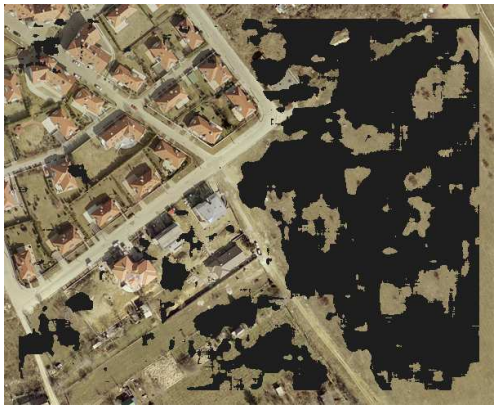
$$b(s) = -\log \int \sqrt{\lambda_s^1(\vartheta) \cdot \lambda_s^2(\vartheta)} d\vartheta$$

The binary similarity map is obtained as $B(s) = 1$ iff $b(s) < b_0$, $B(s) = 0$ otherwise, as shown in Fig. 14.

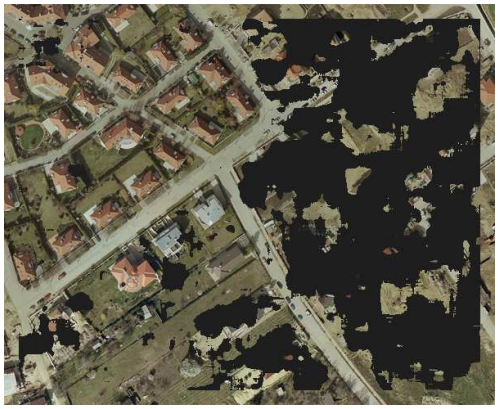
An alternative descriptor for deriving the similarity map is the normalized block correlation. It offers an efficient measure assuming the two regions are identical if and only if the corresponding pixel values are related via an arbitrary linear transform, which is constant for the whole region. Let $\text{mean}_i(s)$ and $\text{var}_i(s)$ be the empirical mean respectively variance values of the gray levels over the $W_l(s)$ subimage of $g^{(i)}$, $i \in \{1, 2\}$ (as before, l is the window size). Derive $\text{corr}(s)$ as the normalized cross correlation coefficient between the

(a) Input image 1 © András Görög

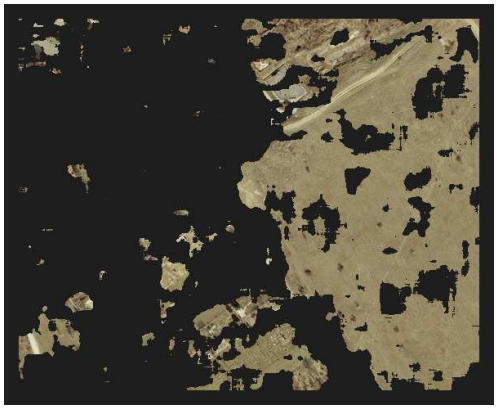
(b) Input image 2 © András Görög



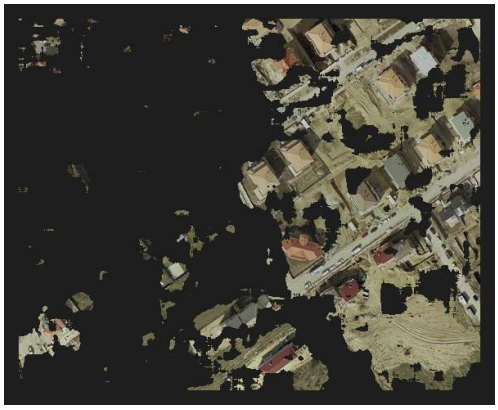
(c) Unchanged regions of image 1



(d) Unchanged regions of image 2



(e) Changed regions of image 1



(f) Changed regions of image 2

Figure 14: Detected changes with the Bhattacharyya distance, BUDAPEST image pair © András Görög. (c) and (d) show the unchanged areas, (e) and (f) present the change mask overlay on images 1 and 2

neighborhoods of s in the two images:[1]

$$\mathrm{corr}(s) = \frac{\sum_{r\in W_l(r)} \big(g_1(r) - \mathrm{mean}_1(s)\big) \cdot \big(g_2(r) - \mathrm{mean}_2(s)\big)}{l^2\sqrt{\mathrm{var}_1(s)\cdot \mathrm{var}_2(s)}} \tag{1}$$

Fig. 13 shows the change detection results with the correlation feature applied for the BEIJING image pair.

3.3 Object-level features

In this section we introduce different object level image features. Based on them we define energy terms denoted by $\varphi^{(i)}(u)$ which evaluate the building hypothesis for u in the i^{th} image (hereafter we ignore again the i superscript). $\varphi(u)$ is interpreted as the negative building fitness value and a rectangle with $\varphi(u) < 0$ is called *attractive* object. Since adding attractive objects may decrease the energy of the population [45], they are efficient building candidates.

We begin with gradient analysis. Below the edges of a good rectangle candidate R_u we expect that the magnitudes of the local gradient vectors are high and the orientations are close to the normal vector of the closest rectangle side (Fig. 15). Λ_u feature is calculated as:

$$\Lambda_u = \frac{1}{q_u} \cdot \sum_{s\in\tilde{\partial}R_u} ||\nabla g_s|| \cdot \big|\cos\big(\vartheta_s - \Theta_u^s\big)\big|$$

where $\tilde{\partial}R_u$ is the edge map of rectangle R_u after dilation, $\Theta_u^s \in \{\theta(u), \theta(u)+90^\circ\}$ is the edge orientation of R_u around $s \in \tilde{\partial}R_u$ and q_u is the number of the pixels in $\tilde{\partial}R_u$. The data-energy term is calculated as: $\varphi_\Lambda(u) = \mathcal{Q}(\Lambda_u, d_\Lambda, D_\Lambda)$ where the following non-linear $\mathcal{Q}$ function is used [45]:

$$\mathcal{Q}(x, d_0, D) = \begin{cases} \left(1 - \frac{x}{d_0}\right) & \text{if } \; x < d_0 \\ \exp\left(-\frac{x-d_0}{D}\right) - 1 & \text{if } \; x \geq d_0 \end{cases}$$

Note that the two parameters of the $\mathcal{Q}$ function can be interpreted easily. Object u is attractive according to the $\varphi_\Lambda(u)$ term iff $\Lambda_u > d_\Lambda$, while D_Λ performs data-normalization before the exponential transform.

The calculation of the *roof color* feature is shown in Fig. 17. We expect that inside the building footprint R_u the image points have dominantly roof colors, while the T_u object-neighborhood (see Fig. 17) should contain background pixels in majority. Hence we calculate the internal $\mathcal{C}_R(u)$, respectively external $\mathcal{C}_o(u)$, filling factors as:

$$\mathcal{C}_R(u) = \frac{1}{\#R_u}\sum_{s\in R_u} \mu_{\mathrm{c}}(s) \qquad \mathcal{C}_o(u) = \frac{1}{\#T_u}\sum_{s\in T_u} \big[1 - \mu_{\mathrm{c}}(s)\big]$$

[1] Using the integral image trick [62] the calculation of the whole correlation map can be efficiently performed.

(a) Object candidate (b) Gradient map (c) Masked gradient map

Figure 15: Utility of the gradient feature

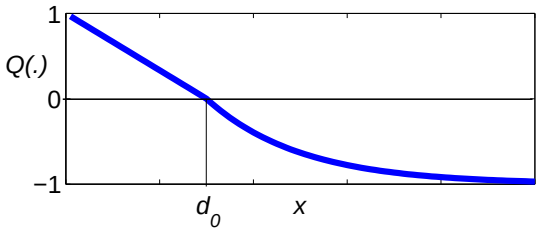


Figure 16: Plot of the $\mathcal{Q}$ function

(a) Red roof (b) Color mask

Figure 17: Utility of the color roof feature

Here $\#X$ denotes the area of X in pixels, and $\mu_c(s)$ is the color mask value in s. We prescribe that u should be attractive according to the color term, if it is attractive both regarding the internal and external subterms, thus the color energy term is obtained as:

$$\varphi_{\mathcal{C}}(u) = \max\left[\mathcal{Q}(\mathcal{C}_R(u), d_R^{\mathcal{C}}, D_R^{\mathcal{C}}), \mathcal{Q}(\mathcal{C}_o(u), d_o^{\mathcal{C}}, D_o^{\mathcal{C}})\right]$$

We continue with the description of the *shadow term*. Using the *shadow direction* vector $\vec{v}_{\text{sh}}$ we identify the two sides of the R_u rectangle which are supposed to border on cast shadows, and denote them by r_1^u and r_2^u (if $\vec{v}_{\text{sh}}$ is parallel with one of the rectangle sides,

Figure 18: Utility of the shadow feature

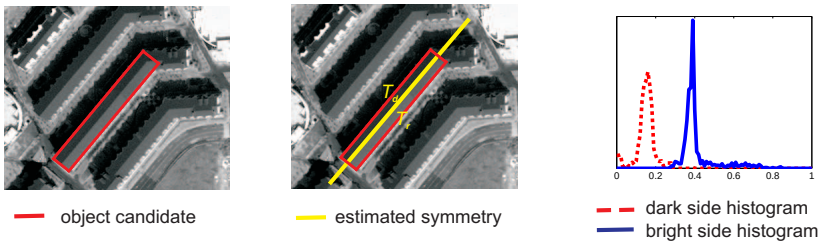


Figure 19: Demonstration of the roof homogeneity feature

we have only one shadow-object edge). For each r_i^u ($i \in \{1,2\}$) we check the presence of shadows in a parallelogram whose parallel sides are r_i^u and $r_i^u + \epsilon_{\text{sh}} \cdot \vec{v}_{\text{sh}}$. Here ϵ_{sh} is a scalar so that $||\epsilon_{\text{sh}} \cdot \vec{v}_{\text{sh}}||$ approximates the shadow width of the shortest buildings in the scene. The union of the r_1^u and r_2^u based parallelograms forms the T_u^{sh} shadow candidate region as shown in Fig. 18. Thereafter, similarly to the color feature, prescribe low shadow presence $\chi_R(u)$ in the R_u internal and high one $\chi_o(u)$ in the T_u^{sh} external region:

$$\chi_R(u) = \frac{1}{\#R_u} \sum_{s \in R_u} \left[1 - \mu_{\text{sh}}(s)\right]; \quad \chi_o(u) = \frac{1}{\#T_u^{\text{sh}}} \sum_{s \in T_u^{\text{sh}}} \mu_{\text{sh}}(s)$$

As for the energy term:

$$\varphi_\chi(u) = \max\left[\mathcal{Q}(\chi_R(u), d_R^\chi, D_R^\chi), \mathcal{Q}(\chi_o(u), d_o^\chi, D_o^\chi)\right]$$

Let us observe that this approach does not require accurate building height information, since we do not penalize if shadow blobs of long buildings exceed the T_u^{sh} regions.

Fig. 19 shows an example of how to describe two-side homogenous roofs. After extracting the symmetry axis of the object candidate u, we can characterize the peakness of the dark, respectively bright, side histograms by calculating their kurtosis, $\kappa_d(u)$ and $\kappa_b(u)$

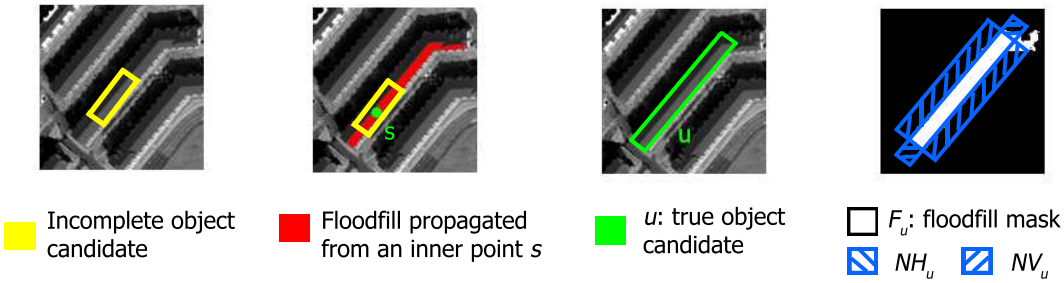


Figure 20: Floodfill based feature for roof completeness

respectively. Denoting the gray value of pixel s by g_s, we get:

$$\kappa_d = \frac{\sum_{T_d} g_s^4}{\left(\sum_{T_d} g_s^2\right)^2} \qquad \kappa_b = \frac{\sum_{T_b} g_s^4}{\left(\sum_{T_b} g_s^2\right)^2}$$

If the roof parts are homogenous, the kurtosis values should be high. However, as Fig. 20 shows the homogeneity feature may have false maxima for incomplete roofs, since parts of a homogenous roof are homogenous as well. Therefore we characterize roof completeness at the same time in the following way. We derive the F_u floodfill mask of u, which contains the pixels reached by floodfill propagations from the internal points of R_u. If the homogenous roof is complete, F_u must have low intersection with the NH_u, resp. NV_u, 'horizontal', and 'vertical', neighborhood regions of R_u (Fig. 20). Finally $\varphi_\kappa(u)$ energy term can be constructed from the kurtosis and completeness descriptors in a similar manner to the previous attributes.

3.3.1 Feature integration

The proposed framework enables flexible *feature integration* depending on the image properties. From the feature primitive terms introduced in Section 3.3, we can construct first building prototypes. For each prototype we can prescribe the fulfillment of one or many feature constraints whose φ-subterms are connected with the max operator in the prototype's joint energy term (logical AND in the negative log-fitness domain).

As well, in a given image pair several building prototypes can be detected simultaneously if the prototype-energies are joined with the min (logical OR) operator. For example in the BUDAPEST pair (see Fig. 24, top) we use two prototypes: the first prescribes the edge and shadow constraints, the second one the roof color alone (as it can detect the red roofs in itself accurately), thus the joint energy is calculated as:

$$\varphi(u) = \min\left\{ \max\{\varphi_\Lambda(u), \varphi_\chi(u)\}, \varphi_c(u)\right\}.$$

Similarly, we used for the BEIJING images (see Fig. 24, bottom) gradient+shadow and homogeneity+shadow prototypes.

4 Marked Point Process model

In this section we transform the building change detection task into an energy minimization problem, which is realized within the Marked Point Process (MPP) framework. For details about MPPs we refer to [63], and for their applications to image processing to [31].

Following our definitions from Section 2, the u building segment candidates (i.e. *objects*) live in a bounded parameter space $\mathcal{H} = \mathcal{C}_x \times \mathcal{C}_y \times \mathcal{E}_L \times \mathcal{E}_l \times \mathcal{O} \times \Xi$. Since we aim to extract building populations from the images, we need to propose a configuration space Ω, which is able to deal with an unknown number of objects:

$$\Omega = \bigcup_{n=0}^{\infty} \Omega_n, \qquad \Omega_n = \big\{\{u_1, \ldots, u_n\} \subset \mathcal{H}^n\big\}$$

Hereafter we will use the notation $\omega \in \Omega$ for an arbitrary object configuration, thus $\omega = \emptyset$, or $\omega = \{u_1, \ldots, u_n\}$ for an $n \in \{1, 2, \ldots\}$ and $u_i \in \mathcal{H} : \forall i \in \{1, 2, \ldots, n\}$.

Emphasizing one of its key features, the MPP framework enables to characterize the whole population instead of individual objects, through exploiting information from entity interactions. Following the classical Markovian approach, each object may only affect its *neighbours* directly. This property limits the number of interactions in the population and results in a compact description of the global scene, which can be analyzed efficiently. To realize the Markov-property, one should define first a $\sim$ neighborhood relation between the objects in $\mathcal{H}$. In our model, we say that $u \sim v$ if their rectangles R_u and R_v intersect.

For characterizing a given ω object population considering the $\mathcal{D}$ image information, we introduce a non-stationary data-dependent Gibbs distribution on the configuration space:

$$P_{\mathcal{D}}(\omega) = \frac{1}{Z} \cdot \exp\Big[-\Phi_{\mathcal{D}}(\omega)\Big]$$

Z being a normalizing constant:

$$Z = \sum_{\omega \in \Omega} \exp\big[-\Phi_{\mathcal{D}}(\omega)\big],$$

and $\Phi_{\mathcal{D}}(\omega)$ a configuration energy:

$$\Phi_{\mathcal{D}}(\omega) = \sum_{u \in \omega} A_{\mathcal{D}}(u) + \gamma \cdot \sum_{\substack{u,v \in \omega \\ u \sim v}} I(u, v) \tag{2}$$

Here $A_{\mathcal{D}}(u)$ and $I\,(u, v)$ are the data dependent unary and the prior interaction potentials, respectively and γ is a weighting factor between the two energy terms. Thus the maximum likelihood configuration estimate according to $P_{\mathcal{D}}(\omega)$ can be obtained as:

$$\omega_{\mathrm{ML}} = \arg\min_{\omega \in \Omega} \big[\Phi_{\mathcal{D}}(\omega)\big]. \tag{3}$$

Unary potentials characterize a given building segment candidate $u = \{c_x, c_y, e_L, e_l, \theta, \xi\}$ as a function of the local image data in both images, but independently of other object of

the population. This term encapsulates the building energies $\varphi^{(1)}(u)$ and $\varphi^{(2)}(u)$ extracted from the 1st, resp. 2nd, image (Sec. 3.3), and the low level similarity information between the two time layers which is described by the $B(.)$ similarity mask (Sec. 3.2).

We remind the reader that our approach attempts to assigns a u object to each building segment, which appears in one $[\xi(u) \in \{1,2\}]$ or both $[\xi(u) = *]$ of the input images. Modified/re-built buildings are considered as two objects u_1 and u_2 corresponding to the appearances in the first and second images respectively, so that $\xi(u_1) = 1$, $\xi(u_2) = 2$.

The following *soft constraints* are considered by the potential terms in the different cases:

- unchanged building u: we expect low object energies in both images, and penalize textural differences under its footprint R_u.
- demolished building or appearance of a changed building in the first image: we expect low $\varphi^{(1)}(u)$, and $\varphi^{(2)}(u)$ is indifferent. We penalize high similarity under the footprint.
- new building or appearance of a changed building in the second image: we expect low $\varphi^{(2}(u)$, and $\varphi^{(1)}(u)$ is indifferent. We penalize high similarity under the footprint.

Consequently, using the $\mathbb{I}_{[E]} \in \{0,1\}$ the indicator function for event E, the $A_{\mathcal{D}}(u)$ potential is calculated as:

$$A_{\mathcal{D}}(u) = \mathbb{I}_{[\xi(u)\in\{1,*\}]} \cdot \varphi^{(1)}(u) + \mathbb{I}_{[\xi(u)\in\{2,*\}]} \cdot \varphi^{(2)}(u) +$$

$$+ \frac{\gamma_\xi}{\#R_u} \left\{ \mathbb{I}_{[\xi(u)=*]} \sum_{s\in R_u} \big(1 - B(s)\big) + \mathbb{I}_{[\xi(u)\in\{1,2\}]} \sum_{s\in R_u} B(s) \right\}$$

On the other hand *interaction* potentials realize prior geometrical constraints: they penalize intersection between different object rectangles sharing the time layer (see Fig. 21):

$$I(u,v) = \mathbb{I}_{[\xi(u)\simeq\xi(v)]} \cdot \frac{\#(R_u \cap R_v)}{\#(R_u \cup R_v)}$$

where $\xi(u) \simeq \xi(v)$ relation holds iff $\xi(u) = \xi(v)$, or $\xi(u) = *$, or $\xi(v) = *$. Note that the intersection term plays a crucial role as it penalizes in (2) to put multiple attractive objects in the same or strongly overlapping positions.

5 Optimization

A recent birth/death dynamics has been proposed by [45], for obtaining an efficient ML approximation of the optimal object configuration (3) in a MPP framework. First the authors defined the dynamics and proved the convergence in continuum, thereafter they proposed a discrete scheme converging to the continuous case. This Multiple Birth and

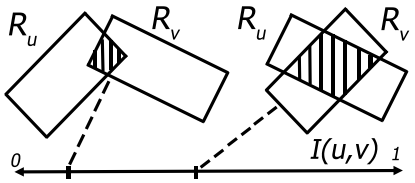


Figure 21: Intersection feature

Table 1: Main properties of the test data sets.

Data Set	Type	Color	Shadow	Gradient	Kurtosis
Budapest	Aerial	Yes	Yes	Good	Partial
Beijing	QBird	No	Yes	Weak	Partial
Szada	Aerial	Yes	No	Weak	No
Abidjan	Ikonos	No	No	Sharp	Yes

Death (MBD) algorithm has been developed to address image processing problems consisting in extracting objects.

The sketch of the MBD algorithm can be followed in Fig. 22: pairs of consecutive birth and death processes are iterated till convergence is obtained in the global configuration. In the *birth* step, multiple object candidates are generated randomly, while *death* attempts to eliminate the inappropriate ones considering the global configuration energy $\varphi(\omega)$.

Therefore the final population - the result of the relaxation after convergence- depends only on the *death* step. However, by using a non homogeneous *birth* rate, the speed of convergence can be significantly increased [45]. As for the parameters of the Multiple Birth and Death optimization process, we followed the guidelines provided in [45]

6 Experiments

We evaluated our method on four significantly different data sets, whose main properties are summarized in Table 1. Qualitative results are shown in Fig. 23–25.

For justification that we have addressed both object extraction and change detection in the same probabilistic framework, we compared the proposed method (hereafter called joint detection - JD) to the conventional approach where the buildings are separately extracted in the two image layers, and the change information is posteriori estimated through comparing the location, geometry and spectral characteristic of the detected objects (separate detection - SD) similarly to [13]. As Fig. 25 shows, the SD method causes false change alarms as low contrasted objects may be erroneously missed from one of the image layers, and due to noise, false objects can appear more frequently in case of less robust one-view information.

Multiple birth and death algorithm for change detection

1. Initialization: calculate the $P_b^{(i)}(s)$, $m_s^{(i)}$ and $\Upsilon_b^{(i)}(s)$ $(i \in \{1,2\})$ birth maps, and start with an empty population $\omega = \emptyset$.

2. Main program: initialize the inverse temperature parameter $\beta = \beta_0$ and the discretization step $\delta = \delta_0$ and alternate birth and death steps:

   - *Birth step*: for each pixel $s \in S$, if there is no object with center s in the current configuration ω, pick up $\xi \in \{1, 2, *\}$ randomly, let be

   $$\widehat{P}_b = \begin{cases} P_b^{(\xi)}(s) & \text{if } \xi \in \{1,2\} \\ \max\{P_b^{(1)}(s), P_b^{(2)}(s)\} & \text{if } \xi = * \end{cases}$$

   and execute the following birth process with probability $\delta \widehat{P}_b$:
   - generate a new object u with center s and image index ξ
   - in case of homogenous roof object
   - set the $e_L(u)$ and $e_l(u)$ side length parameters as follows:
     - if $\Upsilon_b^{(\xi)}(s) = 0$ set the parameters randomly between prescribed maximal and minimal side lengths, following a uniform distribution
     - if $\Upsilon_b^{(\xi)}(s) = 1$ set the parameters according to $\eta\big(., \mu_L^{(\xi)}(s), \sigma_L\big)$ resp. $\eta\big(., \mu_l^{(\xi)}(s), \sigma_l\big)$ Gaussian distributions as explained in Section 3.1.4
   - set the orientation $\theta(u)$ following the $\eta(., m_s^{(\xi)}, \sigma_\theta)$ Gaussian distribution as shown in Sec. 3.1.1
   - add u to the current configuration ω

   - *Death step*: Consider the configuration of objects $\omega = \{u_1, \ldots, u_n\}$ and sort it from the highest to the lowest value of $A(u, \mathcal{D})$. For each object u taken in this order, compute $\Delta\Phi_\omega(u) = \Phi_\mathcal{D}(\omega/\{u\}) - \Phi_\mathcal{D}(\omega)$, derive the *death rate* as follows:

   $$d_\omega(u) = \frac{\delta a_\omega(u)}{1 + \delta a_\omega(u)}, \quad \text{with} \quad a_\omega(u) = e^{-\beta \cdot \Delta\Phi_\omega(u)}$$

   and remove u from ω with probability $d_\omega(u)$. Note that according to (2) $\Delta\Phi_\omega(u)$ depends only on u and its neighbours in ω, thus $d_\omega(u)$ can be calculated locally without computing the global configuration energies $\Phi_\mathcal{D}(\omega/\{u\})$ and $\Phi_\mathcal{D}(\omega)$.

   - *Convergence test*: if the process has not converged, increase the inverse temperature β and decrease the discretization step δ by a geometric scheme and go back to the birth step. The convergence is obtained when all the objects added during the birth step, and only these ones, have been killed during the death step.

Figure 22: Pseudo code of the Multiple Birth and Death (MBD) algorithm

Relevance of the applied multiple feature based building appearance models is compared to the Edge Verification (EV) method. In EV similarly to [3], shadow and roof color information are only used to coarsely detect the built-in areas, while the object verification is purely based on matching the edges of the building candidates to the Canny edge map extracted over the estimated built-in regions.

In the quantitative evaluation we measured the number of missing and falsely detected objects (MO and FO), missing and false change alarms (MC, FC), and the pixel-level accuracy of the detection (DA). For the DA-rate we compared the resulting building footprint masks to the Ground Truth mask, and calculated the F-rate of the detection (harmonic mean of precision and recall). Results in Table 2 confirm the generality of the proposed model and the superiority of the joint detection (JD) framework over the SD and EV approaches (lower object-level errors, and higher DA rates).

7 Conclusion

We have proposed a Marked Point Process framework for building extraction in remotely sensed image pairs taken with significant time differences. The method incorporates object detection and low level change information in a joint probabilistic approach. A global optimization process attempts to find the optimal configuration of buildings, considering the observed data, prior knowledge, and interactions between the neighboring building parts. The accuracy is ensured by a Bayesian object model verification, meanwhile the computational cost is significantly decreased by a non-uniform stochastic object birth process, which proposes relevant objects with higher probability based on low-level image features.

8 Acknowledgement

The first author would like to thank INRIA for funding his work by a twelve-month postdoctoral grant. The authors acknowledge the test data provided by András Görög (Budapest images), French Defense Agency (Abidjan) and Véronique Prinet from LIAMA Laboratory of CAS Beijing (Beijing). As a collaborator in this work, MTA SZTAKI (Hungary) offered the Szada test image pairs.

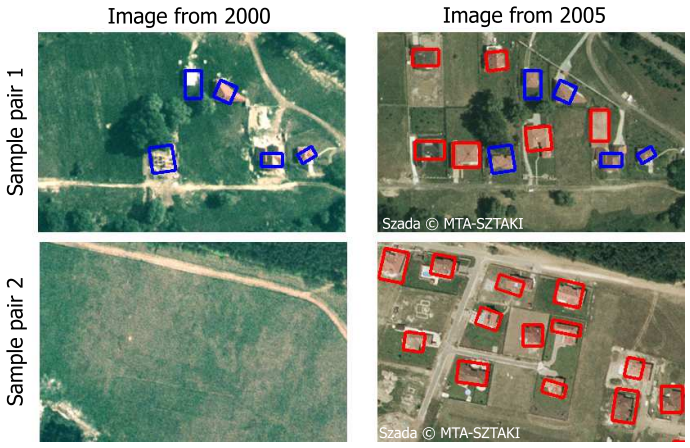


Figure 23: Results on two samples from the SZADA images (source: MTA SZTAKI©). Blue rectangles denote the detected unchanged objects, red rectangles the changed (new, demolished or modified) ones.

Table 2: Quantitative evaluation results. #CH and #UCH denote the total number of changed resp. unchanged buildings in the set. JD (Joint Detection) refers to the proposed model, and for comparison, two reference methods are investigated: EV (Edge Verification) and SD (Separate Detection), see Sec. 6 for details. The object level evaluation rates MO, FO, MC and FC are also defined in Sec. 6.

			MO			FO			MC			FC		
Data Set	#CH	#UCH	EV	SD	**JD**	EV	SD	**JD**	EV	SD	**JD**	EV	SD	**JD**
BUDAPEST	20	21	3	3	1	8	8	2	3	1	1	5	11	1
BEIJING	13	4	0	1	0	5	2	1	0	0	0	2	3	0
SZADA	31	6	4	3	1	1	0	1	3	3	2	2	3	0
ABIDJAN	0	21	1	2	0	0	2	0	0	0	0	0	4	0

Table 3: Quantitative pixel level evaluation results, see Table 2 and Sec. 6 for details.

	DA rate		
Data Set	EV	SD	**JD**
BUDAPEST	0.73	0.70	0.78
BEIJING	0.48	0.77	0.85
SZADA	0.78	0.74	0.83
ABIDJAN	0.84	0.78	0.91

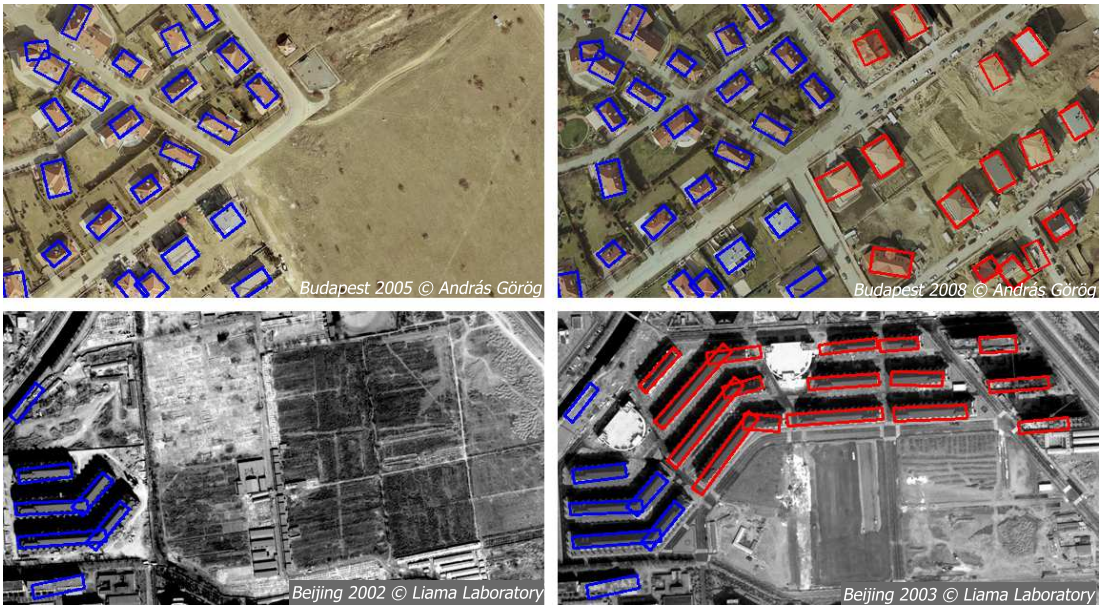


Figure 24: Results on BUDAPEST (top, image part - source: András Görög©) and BEIJING (bottom, LIAMA Laboratory CAS© China) image pairs, marking the unchanged (blue) and changed (red) objects

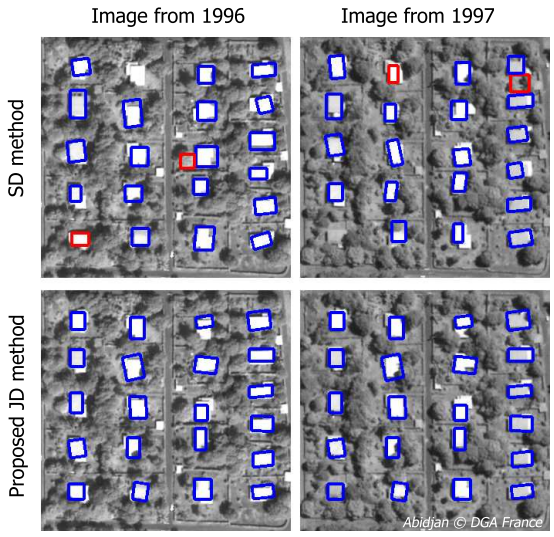


Figure 25: Results on ABIDJAN images (DGA© France). Left: image from 1996, right: image from 1997. Top: separate detection (all changes are false alarms), Bottom: proposed joint model

References

[1] A. Katartzis and H. Sahli. A stochastic framework for the identification of building rooftops using a single remote sensing image. *IEEE Trans. Geosc. Remote Sens.*, 46(1):259–271, 2008.

[2] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny. Structural approach for building reconstruction from a single DSM. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009. To appear.

[3] B. Sirmacek and C. Unsalan. Building detection from aerial imagery using invariant color features and shadow information. In *International Symposium on Computer and Information Sciences (ISCIS)*, Istanbul, Turkey, 2008.

[4] S. Noronha and R. Nevatia. Detection and modeling of buildings from multiple aerial images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(5):501–518, 2001.

[5] F. Bignone, O. Henricsson, P. Fua, and M. Stricker. Automatic extraction of generic house roofs from high resolution aerial imagery. In *European Conference on Computer Vision*, pages I:83–96, Cambridge, UK, 1996.

[6] M. Ortner, X. Descombes, and J. Zerubia. A marked point process of rectangles and segments for automatic analysis of digital elevation models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(1):105–119, 2008.

[7] N. Champion. 2D building change detection from high resolution aerial images and correlation digital surface models. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, page 197, 2007.

[8] D.M. Woo, Q.D. Nguyen, Q.D.N. Tran, D.C. Park, and Y.K. Jung. Building detection and reconstruction from aerial images. In *ISPRS Congress*, Beijing, China, 2008.

[9] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik. Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal for Photogrammetry and Remote Sensing*, 62(2):135–149, 2007.

[10] J.J. Jaw and C.C. Cheng. Building roof reconstruction by fusing laser range data and aerial images. In *Proc. ISPRS Congress*, pages 707–712, Beijing, China, 2008.

[11] J.A. Shufelt. Performance evaluation and analysis of monocular building extraction from aerial imagery. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(4):311–326, 1999.

[12] F. Rottensteiner. Automated updating of building data bases from digital surface models and multi-spectral images: Potential and limitations. In *ISPRS Congress*, Beijing, China, 2008.

[13] S. Tanathong, K.T. Rudahl, and S.E. Goldin. Object oriented change detection of buildings after the Indian ocean tsunami disaster. In *IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pages 65–68, Krabi, Thailand, 2008.

[14] C. Benedek and T. Szirányi. Change detection in optical aerial images by a multi-layer conditional mixed Markov model. *IEEE Trans. Geosc. Remote Sens.*, 47(10):3416–3430, 2009.

[15] A. Fournier, P. Weiss, L. Blanc-Féraud, and G. Aubert. A contrast equalization procedure for change detection algorithms: applications to remotely sensed images of urban areas. In *International Conference on Pattern Recognition (ICPR)*, Tampa, FL, USA, 2008.

[16] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: A systematic survey. *IEEE Trans. on Image Processing*, 14(3):294–307, 2005.

[17] C. Benedek, T. Szirányi, Z. Kato, and J. Zerubia. Detection of object motion regions in aerial image pairs with a multi-layer Markovian model. *IEEE Trans. on Image Processing*, 18(10):2303–2315, 2009.

[18] C. Benedek and T. Szirányi. Bayesian foreground and shadow detection in uncertain frame rate surveillance videos. *IEEE Trans. on Image Processing*, 17(4):608–621, 2008.

[19] L. Bruzzone, D. Fernandez Prieto, and S.B. Serpico. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *IEEE Trans. Geosc. Remote Sens.*, 37(3):1350–1359, 1999.

[20] P. Zhong and R.S. Wang. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. *IEEE Trans. Geosc. Remote Sens.*, 45(12):3978–3988, 2007.

[21] W. Liu and V. Prinet. Probabilistic modeling for structural change inference. In *Proc. Asian Conference on Computer Vision*, pages 836–846, Hyderabad, India, 2006.

[22] L. Castellana, A. D'Addabbo, and G. Pasquariello. A composed supervised/unsupervised approach to improve change detection from remote sensing. *Pattern Recogn. Lett.*, 28(4):405–413, 2007.

[23] C. Benedek and T. Szirányi. Markovian framework for structural change detection with application on detecting built-in changes in airborne images. In *Proc. Int. Conf. on Signal Processing, Pattern Recognition and Applications*, pages 68–73, Innsbruck, Austria, 2007.

[24] L. Bruzzone and S. Serpico. An iterative technique for the detection of land-cover transitions in multitemporal remote sensing images. *IEEE Trans. Geosc. Remote Sens.*, 35(4):858–867, 1997.

[25] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh. A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks. *IEEE Trans. Geosc. Remote Sens.*, 45(3):778–789, 2007.

[26] R. Wiemker. An iterative spectral-spatial Bayesian labeling approach for unsupervised robust change detection on remotely sensed multispectral imagery. In *Proc. Int. Conf. on Computer Analysis of Images and Patterns*, volume LNCS 1296, pages 263–270, Kiel, Germany, 1997.

[27] L. Bruzzone and D. F. Prieto. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *IEEE Trans. on Image Processing*, 11(4):452–466, 2002.

[28] Y. Bazi, L. Bruzzone, and F. Melgani. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans. Geosc. Remote Sens.*, 43(4):874–887, 2005.

[29] L. Bruzzone and D. Fernandez Prieto. Automatic analysis of the difference image for unsupervised change detection. *IEEE Trans. Geosc. Remote Sens.*, 38(3):1171–1182, 2000.

[30] Z.Y. Song, C.H. Pan, Q. Yang, F.X. Li, and W. Li. Building roof detection from a single high-resolution satellite image in dense urban area. In *Proc. ISPRS Congress*, pages 271–277, Beijing, China, 2008.

[31] X. Descombes and J. Zerubia. Marked point processes in image analysis. *IEEE Signal Processing Magazine*, 19(5):77–84, 2002.

[32] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721–741, 1984.

[33] G. Perrin, X. Descombes, and J. Zerubia. A marked point process model for tree crown extraction in plantations. In *Proc. IEEE International Conference on Image Processing (ICIP)*, Genoa, 2005.

[34] G. Perrin, X. Descombes, and J. Zerubia. 2D and 3D vegetation resource parameters assessment using marked point processes. In *Proc. International Conference on Pattern Recognition (ICPR)*, Hong-Kong, 2006.

[35] G. Perrin, X. Descombes, and J. Zerubia. A non-Bayesian model for tree crown extraction using marked point processes. Research Report 5846, INRIA, France, 2006.

[36] S. Descamps, X. Descombes, A. Béchet, and J. Zerubia. Automatic flamingo detection using a multiple birth and death process. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008.

[37] S. Descamps, X. Descombes, A. Béchet, and J. Zerubia. Détection de flamants roses par processus ponctuels marqués pour l'estimation de la taille des populations. Research Report 6328, INRIA France, 2007. In French.

[38] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny. Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. *Journal of Photogrammetry and Remote Sensing*, 63(3):365–381, 2008.

[39] M. Ortner, X. Descombes, and J. Zerubia. Building outline extraction from digital elevation models using marked point processes. *International Journal of Computer Vision*, 72(2):107–132, 2007.

[40] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny. Building reconstruction from a single DEM. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, Anchorage, Alaska, 2008.

[41] L. Garcin, X. Descombes, J. Zerubia, and H. Le Men. Buiding extraction using a Markov point process. In *Proc. IEEE International Conference on Image Processing (ICIP)*, invited paper, Thessaloniki, Greece, 2001.

[42] L. Garcin, X. Descombes, J. Zerubia, and H. Le Men. Building detection by Markov object processes and a MCMC algorithm. Research Report 4206, INRIA, France, 2001.

[43] C. Lacoste, X. Descombes, and J. Zerubia. Point processes for unsupervised line network extraction in remote sensing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1568–1579, 2005.

[44] F. Lafarge, G. Gimel'farb, and X. Descombes. Geometric feature extraction by a multi-marked point process. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009. To appear.

[45] X. Descombes, R. Minlos, and E. Zhizhina. Object extraction using a stochastic birth-and-death dynamics in continuum. *Journal of Mathematical Imaging and Vision*, 33:347–359, 2009.

[46] B. Sirmacek and C. Unsalan. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Trans. Geosc. Remote Sens.*, 47(4):1156–1167, 2009.

[47] P. Saeedi and H. Zwick. Automatic building detection in aerial and satellite images. In *IEEE Intl. Conf. on Control, Automation, Robotics and Vision*, pages 623–629, Hanoi, Vietnam, 2008.

[48] K. Khoshelham and Z.L. Li. A split-and-merge technique for automated reconstruction of roof planes. *Photogrammetric Engineering and Remote Sensing*, 71(7):855–863, 2005.

[49] S. Muller and D.W. Zaum. Robust building detection in aerial images. In *ISPRS Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluation, (CMRT05)*, pages 143–148, Vienna, Austria, 2005.

[50] Z.Y. Song, C.H. Pan, and Q. Yang. A region-based approach to building detection in densely build-up high resolution satellite image. In *Intl. Conference on Image Processing*, pages 3225–3228, Atlanta, Georgia, USA, 2006.

[51] K. Karantzalos and N. Paragios. Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. *IEEE Trans. Geosc. Remote Sens.*, 47(1):133–144, 2009.

[52] J. Peng, D. Zhang, and Y.C. Liu. An improved snake model for building detection from urban aerial images. *Pattern Recognition Letters*, 26(5):587–595, 2005.

[53] C. Benedek and T. Szirányi. Shadow detection in digital images and video. In *Computational Photography: Methods and Applications, Digital Imaging and Computer Vision Book Series*. CRC Press / Taylor & Francis, 2010.

[54] K.L. Chung, Y.R. Lin, and Y.H. Huang. Efficient shadow detection of color aerial images based on successive thresholding scheme. *IEEE Trans. Geosc. Remote Sens.*, 47(2):671–682, 2009.

[55] V.J.D. Tsai. A comparative study on shadow compensation of color aerial images in invariant color models. *IEEE Trans. Geosc. Remote Sens.*, 44(6):1661–1671, 2006.

[56] P.M. Dare. Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering and Remote Sensing*, 71(2):169–178, 2005.

[57] J. Yao and Z.F.M. Zhang. Hierarchical shadow detection for color aerial images. *Computer Vision and Image Understanding*, 102(1):60–69, 2006.

[58] S. Le Hegarat-Mascle and C. Andre. Use of Markov random fields for automatic cloud/shadow detection on high resolution optical images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4):351–366, 2009.

[59] H.T. Guo, Y. Zhang, J. Lu, and G.W. Jin. Research on the building shadow extraction and elimination method. In *ISPRS Congress*, Beijing, China, 2008.

[60] S. Kumar and M. Hebert. Detection in natural images using a causal multiscale random field. In *proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 119–126, Madison, WI, USA, 2003.

[61] *OpenCV documentation.*

[62] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518, Hawaii, USA, 2001.

[63] M. N. M. van Lieshout. *Markov point processes and their applications.* Imperial College Press, London, 2000.

Appendix

A Summary of abbreviations and notations in the report

Abbreviation	**Concept**
MRF	Markov Random Field
KDE	Kernel Density Estimate
MPP	Marked Point Process
MAP	Maximum A Posteriori
ML	Maximum Likelihood
SA	Simulated Annealing (optimization method)
pdf	probability density function

Variable	**Definition**
S	pixel lattice
s, r	pixels ($s, r \in S$)
g_s	gray/intensity value of pixel s
∇g_s	intensity gradient vector at pixel s
$\|\|\nabla g_s\|\|$, ϑ_s	magnitude and angle of the gradient vector
$\eta\,(., m, d)$	Gaussian density function with mean m and standard deviation d
$k(.)$	kernel function
$\lambda_s(\vartheta)$	KDE of local gradient orientations at pixel s
α_s	bi-modal Gaussian similarity feature
$\mathcal{D}$	global image data
$\mathcal{H}$	space of MPP objects
H	a Borel set in $\mathcal{H}$
u, v, $u_i : i = 1 \ldots n$	MPP objects $\in \mathcal{H}$
Ω	configuration space
$\omega = \{u_1, \ldots, u_n\} \in \Omega$	a given object configuration
$\mathcal{V}_u$	neighborhood of object u in ω

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399